논리연구 15-1(2012) pp. 17~52

동전을 던진 후 미녀를 깨우다* **

김 명 석

【요약문】 잠자는 미녀 퍼즐이 등장한 지 벌써 10년이 되었는데 국내 철학자들도 이에 대해 자기 견해들을 내어놓았다. 송하석과 김남중은 미녀의 대답이 1/3이어야 한다고 주장하고 김한승은 관점에 따라 1/2과 1/3이 모두 가능하다고 주장한다. 나는 이 글에서 1/2주의를 선호할 만한 논증을 제안한다. 이를 위해 미녀가 받은 물음이 첫째 물음일 확률은 엘가가 가늠한 것보다 커야 한다는 것을 논증한다. 또한 미녀가 받은 물음이 첫째 물음이라는 것을 그에게 밝혔을 경우, 동전이 앞면이 이미 나왔을 확률이 1/2보다 큰 이유를 해명한다. 하지만 동전 던지는 시점을 미녀가 처음 깨어난 후로 바꿀 경우 오히려 1/3주의 해석이 옳다는 것을 보였다.

【주요어】 잠자는 미녀 문제, 확률, 엘가, 루이스, 송하석

* 접수일자: 2011.09.20. 심사 및 수정 완료일: 2011.11.17. 게재확정일: 2012.01.19.

** 이 논문의 세 심사위원께서는 매우 날카롭고 유익한 논평을 주셨는데 그들의 비평에 답하기 위해 논문을 상당 부분 수정하고 별도의 해명을 새로 첨가해야 했다. 논문을 개선하도록 자극한 이들과 함께, 잠자는 미녀 문제의 매력을 알려준 김한승, 송하석 선생, 애초 내 발표문의 오류를 지적한 선우환 선생, 논문을 읽고 토론해준 생각실험실의 김수민에게 감사드린다.

1. 들어가는 말

잠자는 미녀 퍼즐을 요약하면 다음과 같다.[1] 일요일에 동전을 던진다. 만일 앞면이 나오면 월요일에만 미녀를 깨우고 묻는다. "일요일에 던졌던 동전이 앞면이 나왔을 확률은?" 만일 뒷면이 나오면 월요일에 미녀를 깨우고 묻는다. "일요일에 던졌던 동전이 앞면이 나왔을 확률은?" 그런 다음 질문을 받았다는 기억을 지운 후 잠들게 하고 화요일에 다시 미녀를 깨우고 묻는다. "일요일에 던졌던 동전이 앞면이 나왔을 확률은?" 이러한 설정 전반을 미녀는 알고 있다. 이제 월요일인지 화요일인지 알 수 없는 어느 날, 미녀에게 "일요일에 던졌던 동전이 앞면이 나왔을 확률은?"이라는 물음이 주어진다. 이 때 미녀는 그 확률이 얼마라고 말하는 것이 합당한가?

송하석과 김남중은 미녀의 대답이 1/3이어야 한다고 주장하고 김한승은 관점에 따라 1/2과 1/3이 모두 가능하다고 주장한다.[2] 내가 아는 한, 미녀의 대답이 1/2이어야 한다고 주장하는 국내 철학자의 논문은 아직 등장하지 않았지만, 적지 않은 이들이 1/2이라고 주장하고 있다.[3] 3년 전 나는, 만일 내가 잠자는 미녀이고 확률을 계산할 잠시의 시간이 주어진다면, 나는 "동전이 앞면이 나왔을 확률은 얼마인가?"에 대해 1/3이라고 대답할 것이라고 말한 적이 있다.[4] 하지만 이 글에서 나는 1/3주의가 갖는 문제점을 지적한 후 1/2주의를 옹호하고자 한다. 그러나 동전 던지는 시점을 첫째 물음이 끝난 뒤로 미룰 경우, 1/3주의가 옳다는 것을 보일 것이다.

---

1) Elga (2000); Lewis (2001).

2) 송하석 (2011); 김한승 (2009); 김한승 (2010); 김한승 (2011); Namjoong Kim (2009).

3) 개인 대화에서 과학철학자 이정민과 언어철학자 선우환은 1/2을 지지한다고 말했다.

4) 김명석 (2009), p. 1.

2. 잠자는 미녀의 사후추측

다음과 같은 설정을 생각해 보자. 일요일에 동전을 던진다. 앞면이 나오면 미녀를 깨우지 않는다. 뒷면이 나오면 일주일 중 어느 날 미녀를 깨우고 묻는다. "동전이 앞면이 나왔을 확률은?" 이제 언제인지 알 수 없는 어느 날 미녀에게 "동전이 앞면이 나왔을 확률은?"이라는 물음이 주어진다. 이 때 미녀는 그 확률이 얼마라고 답해야 하는가? 합당한 답은 0일 것이다. 동전이 앞면이 나올 확률은 1/2인데 왜 미녀는 0이라고 답해야 하는가?

동전이 장차 앞면이 나올 확률은 1/2이다. 누군가 이미 던졌던 동전이 앞면이 나왔을 확률도 통상 1/2이다.[5] 이처럼 앞으로 일어날 사건의 확률과 이미 일어난 사건의 확률은 대부분의 경우 같다. 하지만 미래에 벌어질 일을 사전 예측하는 것과 이미 벌어졌지만 자신이 모르는 일을 사후 추측하는 것은 다르다. 그래서 이미 던진 동전이 앞면이 나왔을 확률은, 비록 동전이 멀쩡하고 치우치지 않았다 하더라도, 1/2이 아닐 수 있다. 왜냐하면 이미 벌어진 일을 추측할 때 그 일로부터 정보가 유입될 수 있기 때문이다. 정보는 지각, 보고, 전언 등을 통해 유입될 수 있고, 인과 관계를 통해 추론될 수도 있다. 미래 일은 현재 일에 영향을 줄 수 없지만, 과거 일은 현재 일에 영향을 끼칠 수 있다. 우리는 이 영향을 근거로 과거에 벌어진 일을 추측한다. 미녀의 깨어남은 오직 동전이 뒷면일

5) 던질 동전이 앞면이 나올 확률이 1/2이라는 말은 이를 예측할 만한 정보의 근본 한계를 표현하고 있다. 우리가 아는 한, 동전의 낙하 방향을 결정하는 원인들은 앞면과 뒷면에 대해 완전한 대칭성을 이루고 있다. 다시 말해 동전이 앞면이 나오도록 하는 원인과 뒷면이 나오도록 하는 원인은 균형을 이루고 있다. 우리가 아는 한, 공기와 중력과 바닥 등 모든 원인들과 힘은 앞면과 뒷면 중 어느 쪽을 선호하지 않는다. 더구나 애초의 비대칭성은, 즉 동전이 손을 떠날 때 동전의 특정 배위는 나중에 동전이 어떤 배위를 가지고 바닥에 떨어질지를 예측할 만한 정보를 전혀 제공하지 못한다.

때만 발생한다. 동전이 뒷면이 나오는 사건은 나중에 미녀를 깨어나게 하지만, 앞면이 나오는 사건은 그것을 야기하지 못한다. 그래서 깨어난 미녀는, 자신이 깨어났다는 것을 인지한다면, 또한 그 인지로부터 이러한 인과 추론을 할 수 있다면, "동전이 앞면이 나왔을 확률은?"에 대해 0이라고 답해야 한다.

이와 다른 문제 설정을 생각해 보자. 일요일에 동전을 던진다. 앞면이 나오면 월요일 자정 이후 어느 때 미녀를 깨우고 묻는다. "동전이 앞면이 나왔을 확률은?" 뒷면이 나와도 마찬가지로 미녀를 깨우고 똑같이 묻는다. 언제인지 알 수 없는 어느 때 미녀에게 "동전이 앞면이 나왔을 확률은?"이라는 물음이 주어진다. 이 때 미녀는 그 확률이 얼마라고 답해야 하는가? 미녀는 자기에게 주어진 정보를 모두 동원하여 과거에 무슨 일이 벌어졌는지를 추정할 것이다. 자신이 동원할 정보 가운데 가장 중요한 것은 자신이 지금 깨어나 물음을 받고 있다는 사실이다. 하지만 이 정보는 일요일에 던진 동전이 앞면이 나왔는지 뒷면이 나왔는지를 결정할 만한 정보가 되지 못한다. 달리 말해 앞면과 뒷면에 대한 미녀의 판단에서 완벽한 정보 대칭성이 유지되고 있다.[6] 따라서 미녀는 그 물음에 대해 1/2이라고 답해야 한다.

이제 잠자는 미녀의 애초 문제 설정을 따져 보자. 일요일에 동전을 던진다. 앞면이 나오면 월요일 자정 이후 어느 때 미녀를 깨우고 단 한 번 묻는다. 뒷면이 나오면 미녀를 깨우고 두 번 묻는다. 단 각 물음 이전에 자신이 이미 깨어나 답을 했다는 기억은 제거

6) 우리는 다음과 같은 정보 대칭성 원리 또는 정보 동질성 원리를 가정한다. "동시에 참될 수 없는 두 명제 X와 Y에 대해, 만일 X가 참이라고 판단하게 하는 정보와 Y가 참이라고 판단하게 하는 정보가 똑같다면, 또는 둘 중에 하나를 선호할 만한 정보가 없다면, X가 참일 확률과 Y가 참일 확률은 똑같다." 또는 "동시에 참될 수 없는 두 명제 X와 Y에 대해, 만일 주체가 X와 Y 가운데 무엇이 참인지 판단할 관련 정보를 똑같이 갖고 있다면, X가 참일 확률과 Y가 참일 확률은 그 주체에게 똑같다."

되어야 한다. 이제 언제인지 알 수 없는 어느 때 미녀는 "동전이 앞면이 나왔을 확률은?"이라는 물음을 듣는다.[7] 이 때 미녀는 그 확률이 얼마라고 답해야 하는가? 미녀는 자기에게 주어진 정보를 모두 동원하여 과거에 무슨 일이 벌어졌는지를 추정해야 한다. 미녀가 동원할 정보 가운데 가장 중요한 것은 자신이 지금 깨어나 물음을 받고 있다는 사실이다. 그런데 동전이 앞면이 나오는 사건은 자신을 한 번 깨어나게 하지만, 뒷면이 나오는 사건은 자신을 두 번 깨어나게 한다. 여기에 모종의 비대칭성이 존재하는 것처럼 보인다. 만일 미녀가 앞면과 뒷면의 등장 때문에 자신의 깨어남이 발생했다고 생각한다면, 그리고 자신이 깨어났다는 것을 모종의 정보로 삼는다면, 이 정보는 앞면이 나오는 경우와 뒷면이 나오는 경우에 대해 비대칭 정보로 쓰일 수도 있겠다. 이것은 앞면이 나왔을 확률이 1/2이 아닐 가능성을 열어 놓는다.

잠자는 미녀 문제에 대한 1/3주의자의 답변을 계산하기 위해 다음과 같은 기호들을 도입하자.

Q_n = 월요일 0시 이후 어느 날 미녀가 물음을 받고 있으며 그 물음은 n번째 물음이다.
$Q = Q_1 \vee Q_2 \vee \cdots \vee Q_N$ = 월요일 0시 이후 어느 날 미녀는 질문을 받고 있다.
H = 일요일에 던진 그 동전은 앞면이 나왔다.

---

7) 첫째 물음을 묻는 날이 월요일이어야 한다는 점, 또는 둘째 물음을 묻는 날이 화요일이라는 점은 전혀 중요하지 않다. 둘째 물음을 수요일에 물어도 되고 심지어 월요일 저녁에 물어도 된다. 또한 첫째 물음을 화요일이나 수요일에 물어도 좋다. 심지어 앞면이 나왔을 경우에 수요일에 묻고 뒷면이 나왔을 경우 첫째 물음을 화요일에 물어도 좋다. 잠자는 미녀 문제에서 중요한 것은 앞면이 나오면 물음을 한 번만 묻는다는 것, 뒷면이 나오면 물음을 두 번 묻는다는 것이다. 잠자는 미녀가 깨어난 날이 월요일인지 화요일인지 모른다는 것은 현재 물음이 첫째 물음인지 둘째 물음인지를 모른다는 것이다. 그래서 우리는 물은 날이 월요일인지 화요일인지 따지지 않고, 그 물음이 첫째 물음인지 둘째 물음인지만을 따질 것이다.

T = 일요일에 던진 그 동전은 뒷면이 나왔다.
H_1 = H&Q_1
T_n = T&Q_n
$Pr^s(X)$: s 상황에서 미녀가 X에 대해 가지는 믿음 강도. s 상황이란 미녀가 자신이 깨어나 질문을 받고 있는지 또는 받은 적이 있는지 없는지 모르고 있으며, 동전이 무엇이 나올지 대해 완전한 무지 상태에 있는 상황이다. 대충 말해, 미녀가 일요일에 모든 설명을 듣고, 잠들기 전, 특히 아직 동전을 던지기 전 상황이다.
$Pr^o(X)$: o 상황에서 미녀가 X에 대해 가지는 믿음 강도. o 상황이란 미녀가 자신이 깨어나 지금 질문을 받고 있다는 것을 알지만 그 물음이 몇 번째인지 모르고 있으며, 다만 동전이 이미 던져졌다는 것은 알고 있는 상황이다. 대충 말해, 동전을 이미 던졌으며 미녀가 깨어난 직후 이제 막 물음을 듣고 있는 상황이다. o 상황에서 $Pr^o(X) = Pr^s(X|Q)$. 또한 당연히 $Pr^o(X) = Pr^o(X\&Q)$.
$Pr^+(X)$: + 상황에서 미녀가 X에 대해 가지는 믿음 강도. + 상황이란 미녀가 자신이 깨어나 첫째 질문을 받고 있다는 것을 알고 있으며, 동전이 이미 던져졌다는 것도 알고 있는 상황이다. 대충 말해, 미녀에게 그가 첫째 질문을 받고 있다는 것을 알려주었을 때 상황이다. $Pr^+(X) = Pr^o(X|Q_1)$. 또한 당연히 $Pr^+(X) = Pr^+(X\&Q_1)$.

확률 함수 위에 붙은 첨자 s, o, +는 확률을 계산하는 시점을 표시한 것이지만, 이것을 시점으로 보기보다 미녀가 갖고 있는 정보 상황으로 여기는 것이 낫다.

1/2주의자와 1/3주의자 대부분은 다음에 동의한다.

$$Pr^o(T_1) = Pr^o(T_2)$$

여기서 1/3주의자는 미녀가 지금 물음이 첫째 물음이라는 것을 알게 될 경우, H가 나왔을 확률이 1/2일 것이라고 주장한다.

1/3주의의 근본 가정: $Pr^+(H) = 1/2$.

지금 물음이 첫째 물음이라는 사실은, 이미 나온 동전이 앞면인지 뒷면인지 결정할 만한 아무런 정보도 주지 못한다고 생각했기 때문이다. 1/3주의자는 이로부터 다음을 추론한다.[8)]

$$Pr^{o}(H_1) = Pr^{o}(T_1)$$

이로부터 $Pr^{o}(H_1)$이 1/3이라는 결론을 얻는다. 한편 이들에게 $Pr^{o}(Q_1)$은 2/3이다.

잠자는 미녀 문제에 대한 엘가의 견해를, 루이스의 견해와 대비해, 다음과 같이 정리할 수 있다.[9)]

	$Pr^{s}(H)$	$Pr^{o}(Q_1)$	$Pr^{o}(H)$	$Pr^{+}(H)$
엘가	1/2	2/3	1/3	**1/2**
루이스	1/2	3/4	**1/2**	2/3

여기서 엘가는 $Pr^{+}(H)$를 1/2이라고, 루이스는 $Pr^{o}(H)$를 1/2이라고 애초부터 당연한 것으로 가정한다. 한편 송하석이 잘 지적했듯이, 루이스와 엘가는 둘 다 $Pr^{o}(H)$보다 $Pr^{+}(H)$가 크다는 데 일치를 보고 있다. 그래서 지금 받고 있는 물음이 첫째 물음이라는 것을 아는 순간, 이를 몰랐을 때보다, 앞면이 나올 확률이 커진다는 것에 둘은 동의하고 있다. 하지만 루이스와 엘가의 중요한 차이점은 $Pr^{s}(H)$와 $Pr^{+}(H)$의 크기 차이이다.[10)] 루이스는 후자가 더 크고, 엘가는 두 값이 같다. 왜 미녀는 지금 받고 있는 물음이 첫째 물음이

---

8) $1/2 = Pr^{+}(H) = Pr^{o}(H|Q_1) = Pr^{o}(H\&Q_1)/Pr^{o}(Q_1) = Pr^{o}(H_1)/Pr^{o}(H_1 \vee T_1) = Pr^{o}(H_1)/\{Pr^{o}(H_1) + Pr^{o}(T_1)\}$. 여기서 $Pr^{o}(H_1) + Pr^{o}(T_1) = 2Pr^{o}(H_1)$를 얻는다. 이로부터 $Pr^{o}(H_1) = Pr^{o}(T_1)$.

9) 엘가에게 $Pr^{o}(Q_1) = Pr^{o}(H_1) + Pr^{o}(T_1) = 1/3 + 1/3 = 2/3$이다. 나중에 따로 계산할 텐데, 루이스에게 $Pr^{o}(Q_1) = Pr^{o}(H_1) + Pr^{o}(T_1) = 1/2 + 1/4 = 3/4$이다.

10) 송하석 (2011), p. 9.

라는 것을 아는 순간 앞면이 나왔다는 믿음 강도를 1/2보다 더 높여야 하는가? 송하석은 루이스가 바로 이 물음에 답변해야 한다고 요구했다.

3. 1/3주의의 난점: 너무 작은 $Pr^{o}(Q_1)$

나는 $Pr^{+}(H)$가 1/2보다 커야 한다는 것을 논증함으로써 왜 $Pr^{s}(H)$보다 $Pr^{+}(H)$가 더 큰지를 해명하고자 한다. 이를 위해 먼저 $Pr^{o}(Q_1)$이 엘가가 측정한 것보다 더 큰 값을 가져야 한다는 것을 논증할 것이다. 그 다음 $Pr^{o}(H)$가 1/2이어야 한다는 것을 논증할 것이다. 그러면 잠자는 미녀를 다음과 같은 새로운 문제 설정에 놓아 보자.

일요일에 동전을 던진다. 앞면이 나오면 월요일 자정 이후 어느 때 미녀를 깨우고 묻는다. "동전이 앞면이 나왔을 확률은?" 미녀의 깨어남과 물음은 오직 한 번밖에 이루어지지 않는다. 뒷면이 나오면 월요일 자정 이후 어느 때 미녀를 깨우고 묻는다. "동전이 앞면이 나왔을 확률은?" 그리고 기억을 지우고 재운 다음 깨우고 다시 묻는다. "동전이 앞면이 나왔을 확률은?" 이러한 깨움과 물음을 모두 N번 시행한다. 여기서 N은 2보다 큰 자연수이다. 이제 언제인지 알 수 없는 어느 때 미녀는 "동전이 앞면이 나왔을 확률은?"이라는 물음을 듣는다. 이 때 미녀는 그 확률이 얼마라고 답해야 하는가?[11]

---

11) 이 논문의 한 심사위원은 내 사고실험 설정이 원래 잠자는 미녀 문제가 아니라 다음과 같은 변형된 잠자는 미녀 문제라고 논평했다. "일요일 저녁에 일군의 악당들이 어떤 미녀에게 수면제를 먹인다. 그 후 그들은 각 번호가 나올 확률이 같은 주사위를 던진다. 짝수가 나올 경우—이 가능성을 A라고 하자, 앞뒷면이 나올 확률이 같은 동전을 던진다. 어느 면이 나오건 그들은 미녀를 월요일에 한번만 깨운다. 홀수가 나올 경우—이 가능성을 B라고 하자, 앞뒷면이 나올 확률이 같은 동전을 던지며, 앞면이 나올 경우 그들은 미녀를 월요일에 한번만 깨우지만, 뒷면이 나올 경우 그들은 미녀를 N회

엘가 등 1/3주의자는 다음과 같이 계산할 것이다.

$$Pr^o(T_1) = Pr^o(T_2) = \cdots = Pr^o(T_N)$$
$$Pr^o(H_1) = Pr^o(T_1)^{12)}$$
$$Pr^o(H_1 \vee T_1 \vee T_2 \vee \cdots \vee T_N) = 1$$

이로부터 다음을 얻는다.

$$Pr^o(H_1) = Pr^o(T_1) = Pr^o(T_2) = \cdots = Pr^o(T_N) = 1/(N+1)$$

이제 $Pr^o(Q_n)$를 구해 보자.

$$Pr^o(Q_1) = Pr^o(H_1 \vee T_1) = Pr^o(H_1) + Pr^o(T_1) = 2/(N+1)$$
$$Pr^o(Q_n) = Pr^o(T_n) = 1/(N+1)$$

여기서 n은 2에서 N까지 임의의 자연수이다. 새로운 잠자는 미녀에 대한 엘가의 견해는 다음과 같이 정리할 수 있다.

	$Pr^s(H)$	$Pr^o(Q_1)$	$Pr^o(\sim Q_1)$	$Pr^o(Q_n)$	$Pr^o(H)$	$Pr^+(H)$
2보다 큰 N	1/2	2/(N+1)	(N-1)/(N+1)	1/(N+1)	1/(N+1)	1/2
N = 2 일 때	1/2	2/3	1/3	1/3	1/3	1/2
N이 무한대일 때	1/2	0	1	0	0	1/2

깨우며, 각 깨움 사이에 미녀는 앞에 깨었을 때의 기억을 지우는 주사를 맞는다. 이제 문제는 잠자는 미녀가 깨었을 때에 동전의 앞면이 나올 확률로서 얼마를 할당해야 하는지이다." 솔직히 말해 나의 설정이 왜 이와 같은 변형된 잠자는 미녀 문제와 같은지 파악하는 데 실패했다.

12) 엘가는 $Pr^+(H)$가 1/2라고 가정한다. 이하 계산은 각주 8과 같다.

위 표에서 n은 2에서 N까지 임의의 자연수이다. 1/3주의자의 계산에 따르면, 미녀에게 주어진 물음이 첫째 물음일 확률은 둘째 물음일 확률보다 2배 크다. 하지만 N이 몹시 큰 수일 경우, $Pr^{o}(Q_1)$는 몹시도 작아진다. 별도의 합리적 해명이 없다면, 우리는 이 결과를 받아들이기 어렵다.

N이 몹시 큰 수일 경우 $Pr^{o}(Q_1)$이 몹시도 작아지는 것이 받아들이기 어려운 까닭은 다음과 같다. 미녀에게 물음이 주어지는 방식은 두 가지이다. 하나는 오직 한 번의 물음만 주어지는 방식이다. 다른 하나는 N번의 물음이 주어지는 방식이다. 그래서 나는 다음과 같은 명제를 도입한다.

A = 미녀는 이미 받았던 물음과 앞으로 받을 물음을 모두 합하여 오직 한 번의 질문만 받는다.
B = 미녀는 이미 받았던 물음과 앞으로 받을 물음을 모두 합하여 N번의 질문을 받는다.

여기서 A는 “지금 받고 있는 질문이 첫째 질문이다” 또는 “지금까지 모두 한 번 질문 받았다”가 아님을 유의하기 바란다. A와 B는 서로 배타적이고 $A \vee B$는 참이기 때문에 다음을 얻는다.

$$Pr^{o}(A) + Pr^{o}(B) = Pr^{o}(A \vee B) = 1$$

나는 N이 아무리 크더라도 Pro(A)와 Pro(B) 중 하나가 0으로 수렴되지 않을 것이라 추정한다. 특히 비록 미녀 자신에게 아주 많은 물음이 주어질 가능성이 있다 하더라도, 자신이 바로 지금 물음을 받고 있다는 사실 때문에 명제 A가 거짓이라고 확신하는 것은 마땅해 보이지 않는다.

지금 미녀가 받고 있는 물음이 첫째 물음일 확률을 계산해 보자.

$$Pr^{o}(Q_1) = Pr^{o}(Q_1|A)Pr^{o}(A) + Pr^{o}(Q_1|B)Pr^{o}(B) = Pr^{o}(A) + Pr^{o}(Q_1|B)Pr^{o}(B)$$

여기서 $Pr^{o}(Q_1|A)$가 1이라는 것을 사용했다. 그런데 N이 커질수록 $Pr^{o}(Q_1|B)$도 점차 작아질 텐데, N이 무한대로 접근하면 $Pr^{o}(Q_1|B)$는 0으로 수렴한다. 따라서 $Pr^{o}(Q_1)$는 $Pr^{o}(A)$로 수렴한다. 하지만 엘가 방식의 계산에 따르면, N이 무한대로 커질 때 $Pr^{o}(Q_1)$은 0으로 수렴한다. 결국 다음 네 주장은 동시에 참일 수 없다.

ㄱ. $Pr^{o}(Q_1) = Pr^{o}(A) + Pr^{o}(Q_1|B)(1 - Pr^{o}(A))$
ㄴ. N이 무한대로 커질 때, $Pr^{o}(Q_1|B)$은 0으로 수렴한다.
ㄷ. N이 무한대로 커지더라도, $Pr^{o}(A)$는 0으로 수렴하지 않는다.
ㄹ. N이 무한대로 커질 때, $Pr^{o}(Q_1)$는 0으로 수렴한다.

ㄱ은 조건부 확률의 공리로부터 도출된다. ㄴ은 1/2주의자와 1/3주의자를 포함하여 이 논쟁에 참여하는 거의 모든 이들이 받아들이는 주장이다. 이제 남은 것은 ㄷ과 ㄹ이다. 내 직관은 ㄷ을 유지할 것을 요구한다. 결국 나는 위 네 주장 중에서 ㄹ을 거부한다. 이것이 내가 엘가 방식의 계산을 받아들이지 않는 이유이다. 물론 1/3주의자는 ㄹ을 유지하는 대신에 ㄷ을 거부할 것이다.

1/3주의자의 계산은 $Pr^{o}(\sim Q_1)$을 너무 크게 만들어 버린다. 이 사정을 보다 극단적으로 만들어 보자. 상자에 0부터 100까지 적힌 공 100개가 담겨 있다. 일요일에 공을 꺼내는데 숫자 n이 나오면, 잠자는 미녀를 깨워 다음과 같은 물음을 10^{n}번 묻는다. "일요일에 꺼낸 공에 100이 적혀 있었을 확률은?" 0이 적힌 공이 나오고 물음이 첫째인 경우, 1이 적힌 공이 나오고 물음이 첫째인 경우부터 열째인 경우까지, 2가 적힌 공이 나오고 물음이 첫째인 경우부터 100번째인 경우까지 등 가능한 모든 경우를 세면 $\sum 10^{n} = (10^{101}-1)/(10-1) = (10^{101}-1)/9$이다. 이 수를 M으로 놓자. 1/3주의의 계산에 따르면, 미녀에게 물음이 주어지고 이 물음이 첫째 물음일

확률 $Pr^{o}(Q_1)$은 100/M이다. 왜냐하면 전체 M개 경우들 중에서 첫째 물음인 경우가 모두 100개이기 때문이다. 한편 10^{99}+1에서 10^{100}까지 n에 대해, $Pr^{o}(Q_n)$은 1/M이다. 왜냐하면 이런 물음은 전체 M개 경우들 중에서 하나씩밖에 없기 때문이다. 10^{99}+1번째에서 10^{100}번째까지 물음들의 수는 모두 10^{100}-10^{99} = $9 \cdot 10^{99}$개이다. 1/3주의의 계산에 따르면, 미녀가 지금 받고 있는 물음이 10^{99}번째를 넘긴 물음일 확률은 $9 \cdot 10^{99}/M = 9 \cdot 9 \cdot 10^{99}/(10^{101}-1)$이다. 이 값은 대략 $9 \cdot 9 \cdot 10^{99}/10^{101} = 81/100 = 0.81$이다.

한편 일반 상황에서 상자에서 공을 꺼냈을 때 그 번호가 100일 확률은 0.01밖에 되지 않는다. 반면 그 번호가 100이 아닐 확률은 0.99나 된다. 그런데도 1/3주의의 계산에 따르면, 미녀는 자신이 지금 받고 있는 질문이 10^{99}번째를 넘길 확률이 0.81이나 된다고 생각해야 한다. 하지만 내가 만일 미녀라면, 오히려 지금 주어진 물음이 10^{99}번째를 넘기지 않았을 것이라고 강하게 추정할 것이다. 우리는 공의 개수를 늘여 공에 쓸 숫자를 몹시도 크게 만들 수 있다. 그 수를 N이라고 하자. 1/3주의 계산에 따르면, 미녀에게 주어진 물음이 10^{N-1}번째를 넘길 확률은 0.81이다. 하지만 일반 상황에서 공에 숫자 N이 적혀 있을 확률은 1/N이다. N이 커질수록 이 확률은 줄어들지만, 미녀에게 주어진 물음이 10^{N-1}번째를 넘길 확률은 변함없이 0.81이다. N이 적혀 있는 공이 나왔을 경우, 미녀가 받게 될 질문의 수는 10^{N}인데, 이 수는 미녀가 직면하게 될 경우들의 총수 중 약 90%에 달한다. 1/3주의자는 질문의 수가 이토록 많아지기 때문에, N이 적힌 공이 이미 나왔을 확률이 0.81이나 된다고 판단한 것이다. 하지만 나는 미녀가 10^{N}개의 질문을 받게 되는 세계에 이미 돌입했을 확률이 0.81보다는 훨씬 낮을 것이라고 판단하고 있다. 예컨대 N이 1억이라고 하자. 미녀가 $10^{100000000}$개의 질문을 받는 세계에 돌입할 확률은 0.00000001이다. 하지만 1/3주의

에 따르면, 미녀 자신이 지금 질문을 받고 있다는 사실로부터, 그는 자기가 $10^{100000000}$개의 질문을 받는 세계에 이미 돌입해 있을 확률이 0.81이나 된다고 생각해야 한다. 이러한 반직관적인 결과 때문에 나는 1/3주의를 포기한다.[13]

4. 명백한 논제들

먼저 우리 문제 설정에서 자명한 논제는 다음과 같다. H&B와 T&A는 거짓이다. 그래서

(E1) $Pr^{o}(H|B) = Pr^{o}(B|H) = Pr^{o}(T|A) = Pr^{o}(A|T) = 0$

그 다음 H&T와 A&B는 거짓이고 H∨T와 A∨B는 참이다. 이 사실과 (E1)으로부터 다음을 이끌어낼 수 있다.

(1) $Pro(H|A) = Pro(A|H) = Pro(T|B) = Pro(B|T) = 1$

우리는 이것으로부터 $Pr^{o}(H) = Pr^{o}(A)$를 이끌어낼 수 있다.

---

13) 나의 이 추정은 1/3주의자를 완전히 설득할 만큼 충분히 강하지 않다. 우리 예를 조금 바꾸어, 미녀를 깨워 그냥 질문만 하는 것이 아니라 로또에 당첨되어 하늘을 날 듯 기쁜 기분을 느끼게 한다고 생각해 보자. 미녀는 137억 년 동안 아주 지루하게 잠을 잔다. 그런데 어느 날 깨어나 보니 로또에 당첨되어 하늘을 날 듯 기쁜 기분이 들었다. 미녀는 자신이 처한 이런 상황에 감격하여 자신이 137억 년 동안 $10^{100000000}$번 이런 기분을 느끼는 세계에 이미 들어와 있다고 강하게 믿을지 모르겠다. 하지만 잠자는 미녀의 문제 설정에서 잠에서 깬 미녀는 137억 년이라는 시간 길이와 그 긴 시간의 적막과 지루함을 느끼는 주체가 아니다. 미녀가 의식이 있고 생각하는 모든 시간은 항상 질문 받고 있고 하늘을 날 듯 기쁜 기분이 든다. 이것은 그 일이 오직 한번만 일어나는 세계에서도 마찬가지이다.

(2) $Pr^o(H) = Pr^o(H|A)Pr^o(A) + Pr^o(H|B)Pr^o(B) = Pr^o(A)$
(3) $Pr^o(T) = 1 - Pr^o(H) = 1 - Pr^o(A) = Pr^o(B)$

결국 $Pr^o(H) = Pr^o(A)$는 문제의 설정 상 자명하다.

미녀가 깨어나 지금 질문을 받고 있는 상황에서는 $Q_1 \vee Q_2 \vee \cdots \vee Q_N$는 참이다. 그리고 Q_1&A를 제외한 Q_n&A는 거짓이다. 또한 Q_1&H를 제외한 Q_n&H는 모두 거짓이다. 이 사실들로부터 다음을 얻는다.

(E2) $Pr^o(Q_1 \vee Q_2 \vee \cdots \vee Q_N|A) = Pr^o(Q_1|A) = 1$
(E3) $Pr^o(Q_1 \vee Q_2 \vee \cdots \vee Q_N|B) = 1$
(E4) $Pr^o(Q_1 \vee Q_2 \vee \cdots \vee Q_N|H) = Pr^o(Q_1|H) = 1$
(E5) $Pr^o(Q_1 \vee Q_2 \vee \cdots \vee Q_N|T) = 1$

그리고 H_1&T와 H_1&B와 모든 T_n&H들과 T_n&A들은 거짓이다. Q_1&H와 Q_1&A를 제외하고 모든 Q_n&H, Q_n&A들은 거짓이다. 이 사실로부터 다음을 얻는다.

(E6) $Pr^o(T|H_1) = Pr^o(H_1|T) = Pr^o(B|H_1) = Pr^o(H_1|B) = 0$
(E7) 모든 n에 대해, $Pr^o(H|T_n) = Pr^o(T_n|H) = Pr^o(A|T_n) = Pr^o(T_n|A) = 0$
(E8) $n \geq 2$에 대해, $Pro(H|Qn) = Pro(Qn|H) = Pro(A|Qn) = Pro(Qn|A) = 0$

그리고 $H \vee T$는 참이고 H&T는 거짓이기 때문에, $Pr^o(H|Q_n) + Pr^o(T|Q_n) = 1$, $Pr^o(H|H_1) + Pr^o(T|H_1) = 1$, $Pr^o(H|T_n) + Pr^o(T|T_n) = 1$이다. 또한 $A \vee B$는 참이고 A&B는 거짓이기 때문에 $Pr^o(A|Q_n) + Pr^o(B|Q_n) = Pr^o(A|H_1) + Pr^o(B|H_1) = Pr^o(A|T_n) + Pr^o(B|T_n) = 1$이다. (E6)과 (E7)과 (E8)을 써서

(4) $Pr^o(H|H_1) = Pr^o(A|H_1) = 1$
(5) 모든 n에 대해, $Pr^o(T|T_n) = Pr^o(B|T_n) = 1$
(6) $n \geq 2$에 대해, $Pr^o(T|Q_n) = Pr^o(B|Q_n) = 1$

우리는 (6)으로부터 다음을 얻는다.

(7) $n \geq 2$에 대해, $Pr^o(T_n) = Pr^o(T\&Q_n) = Pr^o(T|Q_n)Pr^o(Q_n) = Pr^o(Q_n)$

하지만 $Pr^o(T_1) = Pr^o(Q_1)$은 성립하지 않는다.
그리고 $H \vee T$는 참이기 때문에 다음은 명백하다.

(E9) $Q_1 \equiv Q_1\&(H \vee T) \equiv (Q_1\&H) \vee (Q_1\&T) = H_1 \vee T_1$

그래서

(8) $Pr^o(Q_1) = Pr^o(H_1 \vee T_1) = Pr^o(H_1) + Pr^o(T_1)$

이제 $n \geq 2$에 대해 다음이 성립한다.

$Pr^o(T_n|B) = Pr^o(T_n\&B)/Pr^o(B) = Pr^o(B|T_n)Pr^o(T_n)/Pr^o(B) = Pr^o(T_n)/Pr^o(B) = Pr^o(Q_n)/Pr^o(B)$

여기서 (5)와 (7)을 사용했다. 또 (6)과 (7)을 사용하여 $n \geq 2$에 대해 다음이 성립한다.

$Pr^o(Q_n|B) = Pr^o(Q_n\&B)/Pr^o(B) = Pr^o(B|Q_n)Pr^o(Q_n)/Pr^o(B) = Pr^o(Q_n)/Pr^o(B)$

이것은 $n \geq 2$일 때 $Pr^o(T_n|B)$와 $Pr^o(Q_n|B)$가 같다는 것을 뜻한다. 그러면 n=1일 때를 고려해 보자. (E9)을 써

$Pr^o(Q_1|B) = Pr^o(H_1 \vee T_1|B) = Pr^o(H_1|B) + Pr^o(T_1|B) = Pr^o(T_1|B)$

를 얻는다. 여기서 $H_1\&T_1$이 거짓이라는 사실과 (E6) 및 (5)를

사용했다. 참고로

$$Pr^o(T_1|B) = Pr^o(T_1\&B)/Pr^o(B) = Pr^o(B|T_1)Pr^o(T_n)/Pr^o(B) = Pr^o(T_1)/Pr^o(B)$$

이다. 결국 $Pr^o(T_1|B) = Pr^o(Q_1|B)$도 성립한다. 나아가 이와 같은 이야기는 $Pr^o(T_n|T)$와 $Pr^o(Q_n|T)$에 대해서 똑같이 할 수 있다. 특히 $Pr^o(T_1|T) = Pr^o(Q_1|T) = Pr^o(T_1)/Pr^o(T)$이며, 또한 (3)에 의해 $Pr^o(T) = Pr^o(B)$이다. 그래서

(9) 모든 n에 대해, $Pr^o(T_n|B) = Pr^o(Q_n|B) = Pr^o(T_n|T) = Pr^o(Q_n|T)$

지금까지 무리한 가정을 단 하나도 하지 않았음을 주목하라.

(E2)에 따르면 $Pr^o(Q_1|A) = 1$이다. (E9)로부터 $Pr^o(H_1 \vee T_1|A) = Pr^o(H_1|A) + Pr^o(T_1|A) = 1$이다. (E7)에 따르면 $Pr^o(T_1|A) = 0$이다. 따라서

(10) $Pr^o(H_1|A) = 1$

(E6)과 (10)로부터 $Pr^o(H_1)$와 $Pr^o(A)$가 같다는 것을 알 수 있다.

(11) $Pr^o(H_1) = Pr^o(H_1|A)Pr^o(A) + Pr^o(H_1|B)Pr^o(B) = Pr^o(A)$

지금까지 놀랄 만한 결과는 전혀 나오지 않았다.[14]

---

14) 이 논문의 한 심사위원은 내가 A와 H가 동치라는 것을 증명 없이 가정하고 있다고 비평했다. 잠자는 미녀가 일요일에 모든 설명을 듣고, 잠들기 전, 특히 아직 동전을 던지기 전에, 그녀도 H이면 A이고, H가 아니면 A가 아니라는 것을 알고 있을 것이다. 미녀가 깨어나 자신이 질문을 받고 있다는 것을 알았을 때, 나아가 지금 받고 있는 질문이 첫째 질문이라는 것을 알았을 때도, 그는 H이면 A이고, H가 아니면 A가 아니라는 것을 알고 있을 것이다. 어느 상황에서도 A와 H가 동치라는 것은 명백한 가정들로부터 논리

(E1)에서 (E9)까지는 모두 자명하며 1/2주의자와 1/3주의자 모두가 받아들일 것이다. 이것들로부터 연역된 (1)부터 (11)까지도 마땅히 받아들여야 한다. 이것들에 덧붙여 나는 다음을 새로 가정한다.

(E10) 모든 n에 대해, $Pr^o(T_1|B) = Pr^o(T_n|B)$

이 가정도 1/2주의자와 1/3주의자가 모두 동의할 것이다. 가정 (E10)과 자명한 논제 (9)로부터 다음이 도출된다.

(12) 모든 n에 대해, $Pr^o(Q_n|B) = Pr^o(T_n|B) = Pr^o(Q_n|T) = Pr^o(T_n|T)$

(E7) 때문에, $Pr^o(T_n) = Pr^o(T_n|A)Pr^o(A) + Pr^o(T_n|B)Pr^o(B) = Pr^o(T_n|B)Pr^o(B)$이다. 그런데 (E10)에 따르면 모든 n에 대해서 $Pr^o(T_n|B)$들이 똑같은 값을 가진다. 이것은 다음을 뜻한다.

(13) $Pr^o(T_1) = Pr^o(T_2) = \cdots = Pr^o(T_N) = Pr^o(Q_2) = \cdots = Pr^o(Q_N)$

여기서 식 (7)을 반영했다. 엘가와 루이스 모두는 이 논제에 동의할 것이다.[15]

그 다음 우리는 자명한 논제 (E3)과 (12)를 이용하여 다음을 얻

적으로 도출된다.

15) 한 심사위원은 나의 문제 설정의 경우 $Pr^o(T_1) > Pr^o(T_2)$이어야 한다고 주장한다. 그의 계산은 다음과 같다. $Pr^o(T_1) = Pr^o(T_1\&(A \vee B)) = Pr^o((T_1\&A) \vee (T_1\&B)) = Pr^o(T_1\&A) + Pr^o(T_1\&B) > Pr^o(T_1\&B) = Pr^o(T_2\&B) = Pr^o(T_2)$. 하지만 나는 여기서 $Pr^o(T_1\&A)$가 0이라고 생각한다. 심사위원이 $Pr^o(T_1\&A)$가 0보다 큰 수라고 생각한 이유는 각주 11에서 언급한 것처럼, 그가 나의 A를 변형된 잠자는 미녀 문제의 A와 동일시했기 때문이다. 하지만 나의 A는, 만일 일단 A가 참이면 T_1은 참일 수 없는 그런 A이다. 나의 A는 "지금 질문이 첫째 질문이다"가 아니며, "지금까지 모두 한 번 질문 받았다"도 아니다. 나의 A는 "지금까지 그리고 앞으로 받을 질문은 모두 합쳐 오직 한 개이다"라는 명제이다.

을 수 있다.

$$1 = Pr^o(Q_1 \vee Q_2 \vee \cdots \vee Q_N|B) = Pr^o(Q_1|B) + Pr^o(Q_2|B) + \cdots + Pr^o(Q_N|B) = NPr^o(Q_1|B)$$

위 식으로부터

(14) 모든 n에 대해, $Pr^o(Q_n|B) = Pr^o(T_n|B) = Pr^o(Q_n|T) = Pr^o(T_n|T) = 1/N$

그리고 (E2)와 (14)로부터 $Pr^o(Q_1)$을 계산해 낼 수 있다.

(15) $Pr^o(Q_1) = Pr^o(Q_1|A)Pr^o(A) + Pr^o(Q_1|B)Pr^o(B) = Pr^o(A) + Pr^o(B)/N$

N을 몹시도 크게 키울 경우 $Pr^o(Q_1)$는 $Pr^o(A)$로 수렴한다.
우리는 (8)과 (11)과 (13)을 사용하여 $Pr^o(T_n)$을 계산해 낼 수 있다.

(16) $Pr^o(T_n) = Pr^o(T_1) = Pr^o(Q_1) - Pr^o(H_1) = Pr^o(A) + Pr^o(B)/N$
$Pr^o(A) = Pr^o(B)/N$

이 식은 1과 N 사이의 모든 자연수 n에 대해 성립한다. 그리고 식 (15)과 (16)을 사용해 $Pr^+(H)$도 간단히 계산해 볼 수 있다.

(17) $Pr^+(H) = Pr^o(H|Q_1) = Pr^o(H\&Q_1)/Pr^o(Q_1) = Pr^o(H_1)/Pr^o(Q_1)$
$= NPr^o(A)/\{NPr^o(A)+Pr^o(B)\}$ [16]

1/2주의자와 1/3주의자 모두는 이 절에서 언급된 모든 논제들을

16) 이 계산은 1/3주의자도 받아들여야 한다. 그들의 계산에 따르면 $Pr^o(A)$는 1/(N+1)이다. 그래서 $Pr^o(B)$는 N/(N+1)이다. 이를 식 (17)에 대입하면 $Pr^+(H) = 1/2$을 얻는다. 달리 말해 $NPr^o(A)/\{NPr^o(A)+Pr^o(B)\}$가 1/2이 되게 하려면 $Pr^o(A)$는 1/(N+1)이 되어야 한다.

받아들여야 할 것이다.

5. 의심의 여지가 있는 논제들

누구나 받아들 수 있는 가정으로부터 우리는 $Pr^s(A) = Pr^s(B) = 1/2$를 이끌어낼 수 있다.

(E11) $Pr^s(H) = Pr^s(T) = 1/2$
(E12) $Pr^s(A|H) = Pr^s(B|T) = 1$
(E13) $Pr^s(A|T) = Pr^s(B|H) = 0$

이 명백한 논제로부터 $Pr^s(A)$를 간단히 계산할 수 있다.

(18) $Pr^s(A) = Pr^s(A|H)Pr^s(H) + Pr^s(A|T)Pr^s(T) = 1/2$
(19) $Pr^s(B) = 1 - Pr^s(A) = 1/2$

$Pr^s(A) = Pr^s(B) = 1/2$이라는 점에 대해서는 아무도 의심하지 않을 것이다.

미녀가 깨어나 지금 질문을 받고 있는 상황에서는 $Q = Q_1 \vee Q_2 \vee \cdots \vee Q_N$는 명백히 참이다. 우리는 $Pr^s(Q)$나 $Pr^s(Q_1)$의 값을 어떻게 할당해야 할까? $Pr^s(Q_1)$ 자체를 정확히 이해하기 어렵지만 여하튼 $Pr^s(Q_1)$의 값을 결정해 보자. 확률 $Pr^s(X)$가 계산되는 상황은 미녀가 자신이 깨어나 질문을 받고 있는지 또는 받은 적이 있는지 없는지 모르고 있으며, 동전이 무엇이 나올지 대해 완전한 무지 상태에 있는 상황이다. 이런 상황에서 $Pr^s(Q_1)$에 값을 준다는 것이 매우 어색하다. 이런 점에서 아래에 나올 논제들은 의심의 여지가 있음을 유념하라.[17] 먼저 일반적으로 다음과 같은 규칙이 성립한다.

---

17) 내가 가정한 논제들 중에서 명백한 것에는 E를 붙였고, 의심의 여지가 있는 것에는 F를 붙였다. (E1)에서 (E13)까지는 명백한 반면, (F1)에서 (F6)까지

$$Pr(A|B\&C) = Pr(A\&B\&C)/Pr(B\&C) = \{Pr(A\&B|C)Pr(C)\}/\{Pr(B|C)Pr(C)\} = Pr(A\&B|C)/Pr(B|C)$$

이 규칙과 논제 (E2)와 식 (14)로부터 다음 논제들을 얻을 수 있다.

(F1) $Pr^s(Q_1|A\&Q) = Pr^s(Q_1\&A|Q)/Pr^s(A|Q) = Pr^o(Q_1\&A)/Pr^o(A) = Pr^o(Q_1|A) = 1$
(F2) $Pr^s(Q_1|B\&Q) = Pr^o(Q_1\&B)/Pr^o(B) = Pr^o(Q_1|B) = 1/N$

나는 이 논제를 s 상황에서 Q, A&Q, B&Q가 거짓이 아니라고 가정하여 도출했다.[18]

나는 다시 의심의 여지가 있는 가정을 도입하고자 한다.

(F3) $Pr^s(Q_1) = Pr^s(Q_1|Q) = Pr^o(Q_1)$

이것은 s 상황에서 o 상황으로 넘어갈 때 Q1의 확률 값이 바뀌지 않는다고 가정하는 것과 같다. 의심의 여지가 있는 논제들 중에서 이것이 나에게 가장 중요하며, 끝내 미결문제로 남게 될 것이다.[19] 나아가 나는 다음 논제를 가정한다. 이 논제 또한 의심의 여지가 있다.

(F4) $Pr^s(Q|A) = Pr^s(Q|B)$

는 의심의 여지가 있다. F류 논제들을 통해 추론된 논제들 역시 의심의 여지가 있지만 별도의 표시는 하지 않았다. 만일 내 주장에 오류가 있다면, 그것은 (F1)에서 (F6)까지 여섯 가정들 가운데서 발견될 것이다.

18) s 상황에서 Q가 거짓이라면 우리는 $Pr^o(X)$를 $Pr^s(X|Q)$로 정의할 수 없을 것이다. $Pr^o(X)$를 $Pr^s(X|Q)$로 정의하면서 우리는 은연중에 $Pr^s(Q)$가 0이 아니라고 가정하고 있다.

19) 혹자는 내가 미결문제의 오류를 범하고 있다고 논평할지 모르겠다. 나는 이것을 받아들인다. 엘가의 미결문제는 "$Pr^+(H) = 1/2$"이고 루이스의 미결문제는 "$Pr^o(H) = 1/2$"이다. 남은 문제는 "$Pr^+(H) = 1/2$", "$Pr^o(H) = 1/2$", "$Pr^s(Q_1) = Pr^s(Q_1|Q)$" 중에서 무엇이 더 직관적인지 판단하는 것이다.

그런데 (F3)으로부터 $Pr^s(Q) = 1$를 이끌어낼 수 있다. Q_1&Q와 Q_1는 동치이기 때문에 $Pr^s(Q_1\&Q) = Pr^s(Q_1)$이다. 따라서 $Pr^s(Q)$가 0이 아닐 경우, $Pr^s(Q_1) = Pr^s(Q_1|Q) = Pr^s(Q_1\&Q)/Pr^s(Q) = Pr^s(Q_1)/Pr^s(Q)$. 결국

(20) $Pr^s(Q) = 1$

이다. 한편 (F4)와 식 (18), (19), (20)을 이용하여 다음을 얻는다.

$$1 = Pr^s(Q) = Pr^s(Q|A)Pr^s(A) + Pr^s(Q|B)Pr^s(B) = 1/2\{Pr^s(Q|A) + Pr^s(Q|B)\} = Pr^s(Q|A)$$

따라서

(21) $Pr^s(Q|A) = Pr^s(Q|B) = 1$

이다.

이제 우리는 확률 $Pr^s(Q_1)$을 계산하고자 한다. 먼저 식 (18), (19), (21)로부터 다음을 얻는다.

(22) $Pr^s(A\&Q) = Pr^s(Q|A)Pr^s(A) = 1/2$
(23) $Pr^s(B\&Q) = Pr^s(Q|B)Pr^s(B) = 1/2$

그 다음 우리는 s 상황에서 다음을 충분히 가정할 수 있다.

(F5) ~Q&A와 ~Q&B가 각각 거짓이거나, $Pr^s(Q_1|\sim Q\&A) = Pr^s(Q_1|\sim Q\&B) = 0$[20]

---

20) 일반 규칙에 따르면, $Pr^s(Q_1|\sim Q\&A) = Pr^s(Q_1\&\sim Q|A)/Pr^s(\sim Q|A)$이다. 여기서 $Pr^s(Q_1\&\sim Q|A)$는 명백히 0이다. 만일 s 상황에서 ~Q&A가 거짓이 아니라면, $Pr^s(\sim Q\&A)$과 $Pr^s(\sim Q|A)$는 0이 아닐 것이다. 당연히 ~Q&A라는 조건이 주어

만일 s 상황에서 ~Q&A와 ~Q&B가 각각 거짓이 아니라면, $Pr^s(Q_1)$은 다음과 같이 쓸 수 있다.

$$Pr^s(Q_1) = Pr^s(Q_1|A\&Q)Pr^s(A\&Q) + Pr^s(Q_1|A\&\sim Q)Pr^s(A\&\sim Q) + Prs(Q1|B\&Q)Prs(B\&Q) + Prs(Q1|B\&\sim Q)Prs(B\&\sim Q)$$

만일 s 상황에서 ~Q&A와 ~Q&B가 모두 거짓일 경우에, $Pr^s(Q_1)$은 조건부 확률의 규칙에 따라 아래와 같이 써야 한다.

$$Pr^s(Q_1) = Pr^s(Q_1|A\&Q)Pr^s(A\&Q) + Pr^s(Q_1|B\&Q)Pr^s(B\&Q)$$

결국 s 상황에서 ~Q&A가 참이든 거짓이든, ~Q&B가 참이든 거짓이든, (F5)에 의해 $Pr^s(Q_1)$은 다음과 같이 쓸 수 있다.

$$(24)\ Pr^s(Q_1) = Pr^s(Q_1|A\&Q)Pr^s(A\&Q) + Pr^s(Q_1|B\&Q)Pr^s(B\&Q) = 1/2 + 1/2N = (N+1)/2N$$

이를 계산하는 데 식 (F1), (F2), (22), (23)이 사용되었다.
이제 (F3)과 식 (24)를 사용하여 마침내 우리는 다음을 얻는다.

$$(25)\ Pr^o(Q_1) = Pr^s(Q_1|Q) = Pr^s(Q_1) = (N+1)/2N$$

식 (25)과 (15)로부터 다음을 얻는다.

$$(26)\ Pr^o(A) + Pr^o(B)/N = (N+1)/2N$$

$Pr^o(B)$은 $1-Pr^o(A)$이기 때문에 (26)로부터

진 상황에서 Q_1이 참일 가능성은 없다. 마찬가지로 $Pr^s(Q_1|\sim Q\&B) = Pr^s(Q_1\&\sim Q|B)/Pr^s(\sim Q|B)$인데 여기서 $Pr^s(Q_1\&\sim Q|B)$는 명백히 0이다. 당연히 ~Q&B라는 조건이 주어진 상황에서 Q_1이 참일 가능성은 없다.

(27) $Pr^o(A) = Pr^o(B) = 1/2$

을 얻는다.[21]

이미 앞에서 $Pr^o(A) = Pr^o(H) = Pr^o(H_1)$이라는 것을 증명했기 때문에

(28) $Pr^o(H) = Pr^o(H_1) = 1/2$

이다. 이것은 루이스의 근본과정과 일치한다. 다른 값들도 쉽게 계산될 수 있다.

(29) $Pr^o(Q_1) = Pr^o(A) + Pr^o(B)/N = (N+1)/2N$
(30) $Pr^o(T_n) = Pr^o(B)/N = 1/2N$
(31) $Pr^+(H) = NPr^o(A)/\{NPr^o(A)+Pr^o(B)\} = N/(N+1)$

N을 몹시도 크게 키울 경우 $Pr^o(Q_1)$, $Pr^o(T_n)$, $Pr^+(H)$는 각각 1/2, 0, 1에 가까워진다. 지금까지 계산한 확률 값들을 간추리면 다음과 같다.

	$Pr^s(H)$	$Pr^o(H)$	$Pr^o(Q_1)$	$Pr^o(Q_n)$	Pr+(H)
2보다 큰 N	1/2	1/2	(N+1)/2N	1/2N	N/(N+1)
N = 2일 때	1/2	1/2	3/4	1/4	2/3
N이 무한대일 때	1/2	1/2	1/2	0	1

21) 식 (20)과 (22)으로부터 곧장 $Pr^o(A) = Pr^s(A|Q) = Pr^s(A\&Q)/Pr^s(Q) = 1/2$을 얻는다. 이 논문의 두 심사위원은 "$Pr^o(A) = 1/2$"의 신빙성에 의문을 제기했다. 한 심사위원은 다음과 같이 말한다. "1/3주의의 직관에 따르자면, 예컨대 동전 앞면이 나오면 1번 깨우고 뒷면이 나오면 100번 깨우는 경우, 잠에서 깨어났을 때 자신에게 던져진 질문이 단 한 번이었을 확률은 1/2보다 훨씬 적다." 1/3주의는 미녀의 깨어남이 A에 대한 믿음 강도를 변경시킨다는 것을 받아들임으로써 "$Pr^o(A) = 1/2$"를 거부한다. $Pr^o(A)$가 1/2보다 작다는 발상은 결국에는 $Pr^o(Q_1)$의 값을 낮추는 역할을 한다. 1/3주의자의 계산에 따르면 N이 커질수록 $Pr^o(Q_1)$은 0에 가까워진다. 하지만 나는 $Pr^o(Q_1)$의 값을 너무 작게 책정해서는 안 된다는 이유에서 1/3주의의 발상을 거부했다.

위 표에서 n은 2에서 N까지 임의의 자연수이다.

6. 몇 가지 교훈들

지금까지 얻은 결과들로부터 중요한 교훈을 얻는다. 무엇보다 미녀 자신이 지금 받고 있는 물음이 첫째 물음일 확률은 항상 1/2보다 크다.

(교훈1) $Pr^{o}(Q_1) > 1/2$

이 값은 N = 2일 때가 가장 크다. N이 커질수록 이 값은 1/2에 가까워진다. 또한 $Pr^{+}(H)$는 언제나 $Pr^{s}(H)$보다 더 크다.

(교훈2) $Pr^{s}(H) < Pr^{+}(H)$

다시 말해 지금 물음이 첫째 물음이라는 정보는 이미 앞면이 나왔음을 뒷받침하는 단서로 쓰일 수 있다. N이 클수록 이 정보는 앞면이 나왔음을 뒷받침하는 결정적인 단서가 된다.

송하석은 1/2주의자가 $Pr^{o}(H)$에 비해 $Pr^{+}(H)$가 더 큰 이유를 설명할 부담을 갖는다고 지적했다.[22] 이에 대해 다음과 같이 변호할 수 있다. 미녀는 자신이 받는 물음이 첫째 물음인지 아닌지 모르는 상태에서 물음을 받았다. 자신이 지금 받고 있는 물음이 둘째일 수 있고 100번째일 수도 천만 번째일 수도 있다. 만일 미녀가 자신이 지금 받고 있는 물음이 첫째 물음이라는 정보를 듣게 된다면, 그는 그토록 많은 가능성들 가운데서 왜 고작 첫째 물음이 주어지고 있

22) 송하석 (2011), p. 9.

는지 깜짝 놀랄 것이다. 이것은 일요일에 뭔가 특별한 일이 일어났다는 것을, 곧 동전이 앞면이 나왔다는 것을 함축한다. 이처럼 묻는 물음이 첫째 물음이라는 정보는, 뒷면이 아니라 앞면이 나왔을 가능성을 높여준다. 이 정보는 뒷면이 나오는 경우와 앞면이 나오는 경우의 정보 대칭성을 깨뜨린다. 결국 미녀는 확신에 차서 일요일에 동전이 앞면이 나왔다고 믿게 될 것이다. 이것이 $Pr^{+}(H)$가 1/2보다 더 커지게 되는 정성적 이유이다.

자신에게 지금 물음이 주어지고 있다는 것을 미녀가 지금 의식하고 있다면, 그는 자신이 오직 한 개의 물음만 주어질 세계에 들어와 있을 가능성은 이제 없어졌다고 생각해야 할까? 그래서 그는 여러 개의 물음이 주어질 세계에 이미 들어와 버렸다고 여겨야 할까? 우리 계산에 따르면 아무리 N이 크더라도, 확률 $Pr^{o}(A)$는 1/2이다. 이것은 자신에게 지금 물음이 주어지고 있다는 사실은 자신이 여러 개의 물음이 주어질 세계에 이미 들어와 버렸다고 판단할 이유가 되지 못한다는 것을 말해준다. 애초에 자신이 그런 세계에 들어올 확률이 여전히 유지된다. 다시 말해

(교훈3) $Pr^{o}(A) = Pr^{s}(A)$

결국 미녀의 깨어남은 A에 대한 믿음 강도를 변경시키지 않는다.

이 상황을 이해하기 위해 다음과 같은 사고실험을 생각해 보자. 마루치 교수는 자정 이전에 동전을 던져 앞면이 나오면 "자정 이전에 던진 동전은 앞면이 나왔는가?"라는 Y형 문제를 하나 출제하여 문제 상자 속에 넣는다. 뒷면이 나오면 N형 문제를 두 개 출제하여 상자 속에 넣는다. 여기서 Y형 문제란 "예"라고 말하면 옳은 답변이 되는 문제이고, N형 문제란 "아니오"라고 말하면 옳은 답변이 되는 문제이다. 미녀는 상자 속에서 "자정 이전에 던진 동전은 앞면이 나왔는가?"라는 물음을 하나 꺼내게 된다. 그런데 미녀

는 상자 속의 모든 문제를 풀어야 한다. 미녀는 상자 속에 문제가 몇 개 있었는지, 몇 개 남았는지 알지 못한다. 상자 속에 문제가 남아 있을 경우 미녀는 약을 먹고 이전에 자신이 문제를 풀었다는 사실을 망각한 채 나머지 문제를 풀어야 한다. 이제 미녀가 받아본 문제가 Y형 문제일 확률은 얼마일까?[23]

---

23) 이 사고실험은 2009년 2월 논리학회 겨울 학술대회에서 발표한 글 "우리는 미녀에게 무엇을 물었을까?"에서 제안한 것이다. 나는 당시 1/3주의를 옹호했다. 송하석 (2011)은 이 발표문에 나오는 나의 문제 설정을 인용했다. 그의 인용은 당시 내 발표문에 비추어 볼 때 정확히 옳았지만, 애석하게도 나는 현재 논문을 쓰면서 견해를 바꾸었다. 이 점을 송 선생께서 양해하시기를 바란다. 나의 애초 견해는 송하석과 거의 일치했다. 특히 그가 이 문제를 접근하는 방식에 대해 나는 완전히 동의했다. "잠자는 미녀의 문제에서 잠자는 미녀에게 물어진 물음은 동전의 본성에 대한 것이 아니라, 이번에 깨어짐이 앞면이 나와서 깨워짐이라고 믿을 합리적인 믿음의 정도에 대한 것이다." 송하석 (2011), p. 15.

아무튼 나는 지금 견해를 바꾸었는데, 그럼에도 불구하고 나의 애초 견해에 중요한 통찰이 담겨 있었다. 내가 당시에 다음과 같이 말한 것은 옳았다. "자정이 지난 후 상자 속에 들어 있는 문제의 수는 하나 아니면 둘이다. 이제 미녀에게 문제 상자가 주어졌다. 이 문제 상자에 오직 Y형 문제만 들어 있을 확률은 1/2이다. 즉 미녀가 상자에서 꺼낸 '오늘 0시에 던진 동전은 앞면이 나왔는가?'라는 물음이 Y형 문제일 확률은 1/2이다. 따라서 이 물음의 답이 Y일 확률은 그녀에게 1/2이다"(일부 중략). 나는 이러한 문제 설정에서 여러 가지 통찰을 이끌어내었는데 지금도 여전히 받아들이고 있는 것은 다음과 같다. 첫째, 미녀의 주관적 확률이 관계하는 개연적 추측은 미래에 벌어질 일에 대한 예측이 아니라 이미 벌어졌지만 자신이 모르는 일에 대한 사후 추측이다. 둘째, 미녀에게 주어질 물음의 출제 방식은 그 물음에 대한 정답을 좌우하는 정보를 포함하고 있다. 그에게 그 출제 방식을 알려주는 것은 그에게 "동전이 앞면이 나왔을 확률은 얼마인가?"의 대답을 좌우할 정보를 준다. 셋째, 미녀가 요구받은 것은 세계 속에서 벌어진 구체적 사건으로서 "동전이 앞면이 나왔을 확률은 무엇인가?"라는 질문자의 발화와 그 전에 동전이 앞면이 나왔을 개연성 사이의 인과적 관련성을 추적하는 것이다. 월요일이 되기 전에 실험 설계에 대한 정보는 주관적 확률을 변화시키지 않는다. 왜냐하면 동전을 던진 결과와 실험 사이에 인과적 연결이 아직 시작되지 않았기 때문이다. 하지만 미녀를 깨우고 물음을 던지는 사건이 일단 발생하면 동전 던지기 사건의 인과적 효과는 시작된다.

하지만 내가 당시에 다음과 같이 말한 것은 잘못되었다. “미녀에게 묻는 시점이 월요일이라는 것을 알려주는 것은 실험 설계를 파괴하는 것 그리하여 그 인과적 효과를 중단시키는 것에 해당한다. 그녀의 주관적 확률은 다시 1/2로 복귀할 수 있다.” 이제 올바르게 말하면, 미녀 자신이 놓여 있는 시점을 알려주는 것은 원래 확률로 복귀시키는 것이 아니라 일어난 일에 대해 새로운 정보를 제공하는 것, 그래서 정보 대칭성을 깨뜨리는 것이다.

나는 당시에 다음 확률이 모두 같다고 판단했다. Y_1 = Y형 물음이 첫째 물음으로 주어진다. N_1 = N형 물음이 첫째 물음으로 주어진다. N_2 = N형 물음이 둘째 물음으로 주어진다. 즉 $Pr^o(Y_1) = Pr^o(N_1) = Pr^o(N_2)$. 그리고 위 세 명제가 다음 세 명제와 각각 동치라고 주장했다. H_1 = 지금 받은 물음은 첫째 물음이고 Y형 물음이다. T_1 = 지금 받은 물음은 첫째 물음이고 N형 물음이다. T_2 = 지금 받은 물음은 둘째 물음이고 N형 물음이다. 당시 발표장에서 선우환 선생께서 이것이 잘못이라는 점을 곧바로 지적했는데 이 점에 대해 뒤늦게 고마움을 전한다. 미녀가 지금 물음을 받고 있는 상황에서, Y_1과 H_1은 논리적 동치이다. “Y형 물음이 첫째 물음으로 주어진다”가 참이라면, 미녀에게 물음이 주어지고 있기 때문에, “지금 받은 물음은 첫째 물음이고 Y형 물음이다”도 참이다. “지금 받은 물음은 첫째 물음이고 Y형 물음이다”가 참이면 당연히 “Y형 물음이 첫째 물음으로 주어진다”도 참이다. 하지만 N_1과 T_1 쌍과 N_2와 T_2 쌍은 서로 동치가 아니다.

또한 나는 당시에 $Pr^o(Y_1) = Pr^o(N_1) = Pr^o(N_2) = 1/2$라고 주장했다. $Pr^o(Y_1)$가 1/2이며, Y_1과 H_1가 우리 상황에서 논리적 동치이며, 그래서 $Pr^o(H_1) = 1/2$이라고 주장한 것을 옳았다. 하지만 $Pr^o(Y_1) + Pr^o(N_1) + Pr^o(N_2)$가 1.5가 된다는 사실 때문에 나는 이 확률의 합이 1이 되도록 각 확률들을 재규격화해야 했다. 이를 위해 우리의 시점과 다른 새로운 시점 즉 “미녀가 깨었을 때”라는 새로운 시점을 도입했다. 당시 나는 다음과 같이 주장했다. “망각제는 월요일과 화요일의 시간 간격을 제거하는 효과를 준다. 따라서 망각제는 실제 세계에서 월요일 시점에 발생 가능한 두 사건과 화요일 시점에 발생 가능한 한 사건이 마치 동일한 시간에 발생 가능한 세 사건인 것으로 착각하게 만든다. 미녀에게 그 세 사건은 가상적 세계, 망각제에 의해 형성된 그녀 자신의 시공간 내 어느 한 시점에 벌어질 만한 세 사건이다. 즉 그녀에게는 그 세 사건이 동일한 표본공간에 속하는 사건으로 간주된다. 표본공간 내 사건 발생 확률의 총합은 1이 되어야 하기 때문에 이 세 사건이 발생할 확률의 합계는 1.5가 아니라 1로 재조정되어야 한다.” 하지만 H_1, T_1, T_2는 이러한 새로운 시점을 도입할 필요가 없고 각 명제들 속에 이미 그러한 시점이 반영되어 있다. 나는 확률을 재규격화하면서, $Pr^o(Y_1) = Pr^o(N_1) = Pr^o(N_2)$를 $Pr^o(H_1) = Pr^o(T_1) = Pr^o(T_2)$이 되도록 했다. 결국 나는

미녀는 지금 깨어나 문제를 하나 받게 된다. 자신이 지금 문제를 풀고 있다는 사실은 그가 풀어야 문제의 개수가 하나인지 둘인지 판단할 아무런 정보를 주지 못한다. 자신은 이미 Y형 문제를 하나 푸는 A 세계와 N형 문제를 둘 푸는 B 세계 중 하나의 세계에 이미 돌입했다. 자신이 A 세계에 돌입했을 확률은 애초에 A 세계에 들어갈 확률과 같다고 여길 것이다. 자신이 앞으로 A 세계에 들어갈 확률과 B 세계에 들어갈 확률이 같다면, 자신이 이미 A 세계에 들어왔을 확률과 B 세계에 들어왔을 확률이 같다고 생각할 것이다. 하지만 1/3주의자에 따르면, 지금 물음이 물어지고 있다는 사실은

$Pr^o(H_1) = Pr^o(T_1) = Pr^o(T_2) = 1/3$이라고 주장했는데 이것은 잘못된 결론이다.

이 모든 오류는 내가 $N_1 \equiv T_1$과 $N_2 \equiv T_2$를 섣불리 가정했기 때문이다. 당시에 나는 다음과 같이 잘못 주장했다. "[미녀의] 가상적 세계에서 H_1, T_1, T_2가 발생할 확률은 1/2에서 1/3로 재조정되어야 한다. 이것은 미녀의 착각일 뿐 실제 세계 즉 우리의 시공간에서 H_1, T_1, T_2가 발생할 확률은 여전히 1/2이다." 바로 잡아 말하자면, 우리의 시공간에서 Y_1, N_1, N_2가 발생할 확률은 1/2이고 이것은 미녀에게도 마찬가지이다. 나는 당시에 이미 미녀의 시공간에서 N_1와 N_2가 논리적으로 동치라고 주장했었다. (지금 생각해 보면 사실 우리의 시공간에서도 둘은 동치이다.) 논리적 동치인 명제 P와 Q에 대해, $Pr(P \vee Q) = 2Pr(P)$라고 계산하는 것은 완벽한 오류였다.

분명 N_1과 T_1은 동치가 아니다. N_1이 참이라면, 일요일에 던진 동전이 뒷면이 나왔다는 것을 뜻한다. 미녀에게는 첫째 물음과 둘째 물음이 모두 반드시 주어질 것이다. 하지만 미녀는 지금 받고 있는 물음이 첫째 물음인지 둘째 물음인지 알 수 없다. 그래서 T_1이 참이라고 말할 수 없다. 반면에 T_1이 참이라면 N_1은 참이다. 또한 T_2이 참이라면, N_2는 참이다. 물론 N_1가 참이면 N_2도 반드시 참이고 그 역도 성립한다. 간추려 말하면, $T_1 \Rightarrow N_1$, $T_2 \Rightarrow N_2$, $N_1 \Leftrightarrow N_2$. 잘 알려져 있다시피, 명제 X가 Y를 함축한다면 $Pr(X) \leq Pr(Y)$를 만족한다. 여기서 등호는 둘이 논리적 동치일 때 성립한다. 따라서 우리는 다음과 같이 주장할 수 있다. $Pr^o(T_1) < Pr^o(N_1) = Pr^o(N_2) = 1/2$. $Pr^o(T_2) < Pr^o(N_1) = Pr^o(N_2) = 1/2$. 이처럼 내가 보다 신중했다면 $Pr^o(T_1)$와 $Pr^o(T_2)$가 1/2보다 작아야 한다고 결론 내렸을 것이다. 한편 N_1은 '$T_1 \vee T_2$'를 함축하고 그 역도 성립한다. 따라서 $1/2 = Pr^o(N_1) = Pr^o(T_1 \vee T_2) = Pr^o(T_1) + Pr^o(T_2)$. $Pr^o(T_1)$와 $Pr^o(T_2)$가 같다는 가정으로부터 $Pr^o(T_1)$가 1/4라는 것을 쉽게 얻어낼 수 있다.

자신이 물음이 물어지는 사건이 많이 벌어지는 세계에 이미 들어가 있다는 믿음을 강화해준다. 미녀가 지금 문제를 풀고 있다면, 문제 수가 많은 세계에 자신이 들어와 있을 가능성이 더 높다고 판단해야 한다는 것이다.

7. 미녀를 깨운 후 동전을 던지다.

엘가와 루이스의 경우, 월요일에 미녀를 깨우고 질문하고 재운 뒤에 비로소 동전을 던져도 자신들의 결론이 동일하다고 주장하는 것으로 알려져 있다.[24] 이 경우에 나의 확률 계산은 다소 의외의 결과가 나왔다. "월요일 첫째 질문에 답한 뒤 던질 동전은 앞면이 나온다"를 H*이라고 하자. 내 생각에 $Pr^{+}(H^*)$은 1/2이다. 이것은 우리 문제를 푸는 실마리이다.

(E14) $Pr^{+}(H^*) = 1/2$

---

24) 이 논문의 한 심사위원은 이 논문의 결함을 지적하면서 동전을 던지는 시점에 대한 나의 설정이 원래 설정에 비해 제한적이라고 논평했다. 그는 엘가와 루이스의 다음 구절을 인용해주셨다. "They might accomplish their task by either (1) first tossing the coin and then waking you up either once or twice depending on the outcome; or (2) first waking you up once, and then tossing the coin to determine whether to wake you up a second time. Your credence (upon awakening) in the coin's landing Heads ought to be the same regardless of whether the researchers use method (1) or (2). So without loss of generality suppose that they use－and you know that they use－method (2)." Elga (2000), pp. 144-145. "I haven't said yet whether the coin was to be tossed before or after the Monday awakening. Elga's argument applies in the first instance to the case that it is tossed after; but he thinks, and I agree, that the answer to our question should be the same in both case." Lewis (2001), p. 172. 동전 던지는 시점을 변경할 경우, 당혹스럽게도, 나의 확률이 변경된다는 것을 논문을 수정하면서 발견하였다. 심사위원의 물음에 답하면서 여기 새로운 절을 추가했는데 이 절에서 기술된 나의 계산이 옳기를 바란다. 옳든 그르든 그에게 감사를 표한다.

여기서 $Pr^{+}(X)$는 이미 정의한 대로 미녀가 자신이 깨어나 첫째 질문을 받고 있다는 것을 알고 있으며, 동전이 이미 던져졌다는 것도 알고 있는 상황에서 X의 믿음 강도이다. 자신이 받고 있는 물음이 첫째 물음이라는 것을 아는 순간 미녀는 아직 우리가 동전을 던지지 않았다는 것도 알게 된다. 만일 엘가가 $Pr^{+}(H^*) = 1/2$라고 주장한다면, 그는 옳다. 또한 루이스가 $Pr^{+}(H^*) = 1/2$라고 주장한다면, 그는 옳다. 하지만 만일 루이스가 $Pr^{+}(H^*) = 2/3$라고 주장한다면, 그는 틀렸다. 다른 확률 값들을 계산하기 위해 몇몇 명제들을 재정의하자.

Q^*_1 = 동전을 아직 던지지 않았지만 월요일 이후 어느 날 미녀가 물음을 받고 있으며 이 물음은 첫째 물음이다.
$n \geq 2$에 대해, Q_n = 동전을 이미 던졌고 월요일 이후 어느 날 미녀가 물음을 받고 있으며 이 물음은 n번째 물음이다.
$Q^* = Q^*_1 \vee Q_2 \vee \cdots \vee Q_N$
H^* = 월요일에 던질 또는 던진 동전이 앞면이 나온다.
T^* = 월요일에 던질 또는 던진 동전이 뒷면이 나온다.
$n \geq 2$에 대해, $T^*_n = T^* \& Q_n$
$Pr^{o}(X) = Pr^{s}(X|Q^*) = Pr^{o}(X \& Q^*)$
$Pr^{+}(X) = Pr^{o}(X|Q^*_1) = Pr^{+}(X \& Q^*_1)$

제4절에서 증명된 명백한 논제들도 거의 그대로 유지된다.

(2*) $Pr^{o}(H^*) = Pr^{o}(A)$
(3*) $Pr^{o}(T^*) = = Pr^{o}(B) = 1 - Pr^{o}(A)$
(E3*) $Pr^{o}(Q^*_1 \vee Q_2 \vee \cdots \vee Q_N|B) = 1$
(E8*) $n \geq 2$에 대해, $Pr^{o}(H^*|Q_n) = Pr^{o}(Q_n|H^*) = Pr^{o}(A|Q_n) = Pr^{o}(Q_n|A) = 0$
(6*) $n \geq 2$에 대해, $Pr^{o}(T^*|Q_n) = Pr^{o}(B|Q_n) = 1$
(7*) $n \geq 2$에 대해, $Pr^{o}(T^*_n) = Pr^{o}(Q_n)$
(9*) $n \geq 2$에 대해, $Pr^{o}(T^*_n|B) = Pr^{o}(Q_n|B) = Pr^{o}(T^*_n|T^*) = Pr^{o}(Q_n|T^*)$

(9*)는 (9)와 달리 일단은 n=1인 경우를 제외했다. 이제 1/2주의자와 1/3주의자 모두가 받아들일 다음 논제를 가정한다. 역시 n=1

인 경우를 제외했다.

(E10*) $n \geq 2$에 대해, $Pr^{o}(T^{*}_{2}|B) = Pr^{o}(T^{*}_{n}|B)$

이 가정과 위 논제들로부터 다음이 도출된다. 역시 n=1인 경우가 빠졌다.

(12*) $n \geq 2$에 대해, $Pr^{o}(Q^{*}_{2}|B) = Pr^{o}(Q_{n}|B)$
(13*) $Pr^{o}(T^{*}_{2}) = Pr^{o}(Q^{*}_{2}) = Pr^{o}(T^{*}_{3}) = Pr^{o}(Q^{*}_{3}) = \cdots = Pr^{o}(T^{*}_{N}) = Pr^{o}(Q^{*}_{N})$

이제 다음을 새로 가정해야 한다.

(E14) $Pr^{+}(H^{*}) = Pr^{+}(A^{*}) = Pr^{+}(T^{*}) = Pr^{+}(B^{*}) = 1/2$

이 값들이 1/2인 이유는 이들 확률이 사후 추측이기 때문이다. (E14)는 거의 의심의 여지가 없다.

먼저 (E14)로부터 우리는 다음을 얻을 수 있다.

(32) $Pr^{o}(H|Q^{*}_{1}) = Pr^{o}(A|Q^{*}_{1}) = Pr^{o}(T|Q^{*}_{1}) = Pr^{o}(B|Q^{*}_{1}) = 1/2$

당연히 (E8*)과 (6*)에 따르면 $Pr^{o}(A|{\sim}Q^{*}_{1}) = 0$이고, $Pr^{o}(B|{\sim}Q^{*}_{1}) = 1$이기 때문에 다음을 얻는다.

(33) $Pr^{o}(A) = Pr^{o}(A|Q^{*}_{1})Pr^{o}(Q^{*}_{1}) + Pr^{o}(A|{\sim}Q^{*}_{1})Pr^{o}({\sim}Q^{*}_{1}) = Pr^{o}(Q^{*}_{1})/2$
(34) $Pr^{o}(B) = Pr^{o}(B|Q^{*}_{1})Pr^{o}(Q^{*}_{1}) + Pr^{o}(B|{\sim}Q^{*}_{1})Pr^{o}({\sim}Q^{*}_{1}) = Pr^{o}(Q^{*}_{1})/2 + Pr^{o}({\sim}Q^{*}_{1})$

(33)과 (34)를 잘 비교해 보면 $Pr^{o}(B)$보다 $Pr^{o}(A)$가 더 작다는 것을 알 수 있다.

(35) $Pr^{o}(A) < Pr^{o}(B)$

여기서 등호가 빠진 이유는 확률 $Pr^{o}(\sim Q^{*}_{1})$이 0이 아니기 때문이다. 왜 미녀는 $Pr^{o}(A)$와 $Pr^{o}(B)$가 같은 값을 가진다고 생각하지 않아야 하는가? 만일 자신이 A 세계나 B 세계 중 어느 하나에 아직 돌입하지 않았다고 생각한다면, 그는 두 확률이 같다고 여길 것이다. 하지만 만일 자신이 이미 두 세계 중 하나에 이미 돌입했다고 생각한다면, 이것은 지금 자신에게 주어진 물음이 첫째 물음이 아니라는 것을 뜻하며, 결국 그는 자신은 이미 B 세계에 돌입해 있다고 확신해야 한다. 이러한 비대칭성 때문에 그는 자신이 B 세계에 들어와 있을 확률이 더 높다고 판단해야 한다.[25]

식 (32)과 (33)과 (13*)로부터 다음을 얻는다.

(36) $Pr^{o}(Q^{*}_{1}) = 2Pr^{o}(A)$
(37) $Pr^{o}(Q^{*}_{2}) = Pr^{o}(\sim Q^{*}_{1})/(N-1) = \{1-2Pr^{o}(A)\}/(N-1)$
(38) $Pr^{o}(Q^{*}_{2}|B) = \{1-2Pr^{o}(A)\}/\{(1-Pr^{o}(A))(N-1)\}$

식 (38)은 식 (6*)로부터 $Pr^{o}(Q^{*}_{2}|B) = Pr^{o}(Q^{*}_{2})/Pr^{o}(B)$라는 사실을 이용해 얻었다. 지금까지 계산을 도표로 정리하면 다음과 같다.

	$Pr^{o}(Q^{*}_{1})$	$Pr^{o}(\sim Q^{*}_{1})$	$Pr^{o}(H^{*})$	$Pr^{o}(T^{*}_{2})$	$Pr^{o}(T^{*})$	$Pr^{+}(H^{*})$
모든 N	$2Pr^{o}(A)$	$1-2Pr^{o}(A)$	$Pr^{o}(A)$	$\{1-2Pr^{o}(A)\}/(N-1)$	$1-Pr^{o}(A)$	1/2

이 결과는 1/2주의자와 1/3주의자 모두가 받아들일 것이다.

나는 의도적으로 (E10*)에서 T_1을 생략했다. 이렇게 한 이유는 이 명제가 나에게 다소 모호했기 때문이다. 1/2주의자와 1/3주의자 모두 (E10*)에서 T_1을 포함하는 것을 주저하지 않는다. 비록 다소

25) 하지만 월요일 아침 이전에 동전을 이미 던졌다면, 자신은 A 세계와 B 세계 중에서 이미 하나에 돌입해 있다고 판단하겠지만, 그는 이것으로부터 곧 자신에게 주어진 물음이 첫째 물음이 아니라고 추론하지 않아야 한다.

의심의 여지가 있지만, 나도 다음 논제를 가정한다.

(F6) $Pr^{o}(T^{*}_{1}|B) = Pr^{o}(T^{*}_{2}|B)$

이 논문에서 이 가정도 일단 미결문제로 남겨 놓는다. 여하튼 이를 가정할 경우 다음도 받아들여야 한다.

(39) $Pr^{o}(Q^{*}_{1}|B) = Pr^{o}(Q^{*}_{2}|B)$

이것은 식 (12*)에 n=1인 경우를 포함시키는 것이다. (39)로부터 $Pr^{o}(A) = 1/(N+1)$을 계산해낼 수 있다. 먼저 (E3*)와 (F6)로부터

$1 = Pr^{o}(Q^{*}_{1} \vee Q_{2} \vee \cdots \vee Q_{N}|B) = Pr^{o}(Q^{*}_{1}|B) + Pr^{o}(Q_{2}|B) + \cdots + Pr^{o}(Q_{N}|B) = NPr^{o}(Q^{*}_{1}|B)$

를 얻는다. 그리고 (38)과 (39)로부터

$1/N = \{1-2Pr^{o}(A)\}/\{(1-Pr^{o}(A))(N-1)\}$

을 얻는다. 이로부터 다음을 얻는다.

(40) $Pr^{o}(A) = 1/(N+1)$
(41) $Pr^{o}(B) = 1-Pr^{o}(A) = N/(N+1)$

이를 도표로 정리하면 다음과 같다.

	$Pr^{o}(Q^{*}_{1})$	$Pr^{o}(\sim Q^{*}_{1})$	$Pr^{o}(H^{*})$	$Pr^{o}(T^{*}_{2})$	$Pr^{o}(T^{*})$	$Pr^{+}(H^{*})$
모든 N	2/(N+1)	(N-1)/(N+1)	1/(N+1)	1/(N+1)	N/(N+1)	1/2
N = 2	2/3	1/3	1/3	1/3	2/3	1/2
$N \rightarrow \infty$	0	1	0	0	1	1/2

이것은 1/3주의자의 결과와 정확히 일치한다. 결국 우리는 동전 던지는 시점이 잠자는 미녀 문제에서 중요하다고 결론 내려야 한다. 미녀가 처음 깨어나기 이전에 동전을 던지느냐, 아니면 첫째 물음을 마친 뒤에 동전을 던지느냐 하는 것은 동전이 앞면이 나올 확률 또는 나왔을 확률을 변경시킨다. 두 사고실험 설정은 똑같은 결과를 낳지 않는다.

그렇다면 첫째 물음 후에 동전을 던지는 상황이 왜 1/3주의 해석과 같아야 하는가? 미녀가 직면하게 될 총 N+1개의 경우들 중에 오직 한 경우만이 A 세계와 B 세계 중 선택이 아직 열려 있다. 결정이 열려 있을 경우 1/2 확률을 지닌 채 각 세계에 들어가게 될 것이다. 이 경우에 H와 T 사이에 대칭성이 유지된다. 하지만 N+1개 경우들 중 N개 경우들은 이미 선택이 닫혀 있으며 더구나 N개 모두 B 세계로 이미 결정 나 있다. 결국 미녀는 N/(N+1) 확률로 이미 B 세계에 들어와 있다고 믿는 것이 옳다. 다시 말해 N이 몹시 클 경우 자신이 A 세계에 돌입해 있을 가능성은 거의 없다. 이것은 1/3주의자가 잠자는 미녀 문제를 해석하는 방식과 정확히 일치한다. 하지만 미녀가 처음 깨어나기 이전에 이미 동전을 던졌다면 자신은 두 세계 중 어느 하나에 이미 돌입해 있다고 믿는 것이 옳다. 그래서 아무리 N이 크더라도 자신이 A 세계에 이미 돌입해 있을 가능성이 0으로 낮아져서는 안 된다. 이런 이유 때문에 나는 1/3주의를 포기했다.

8. 나오는 말

송하석과 김한승은 “두 아이의 역설”이 역설이라고 생각한다.[26] 김한승은 잠자는 미녀 문제가 역설이라고 생각하지만 송하석은 그렇

26) 김한승 (2009), pp. 127-129; 송하석 (2011), p. 17.

지 않다. 우리가 마주하는 문제들이 우리를 당혹케 하는 것은 사실이지만, 나는 대부분의 문제들이 역설이 아니라 일종의 수수께끼라고 생각한다. 이들의 작업들이 잘 보여주듯이, 역설과 수수께끼 속에 담긴 인간 인식과 이성의 미로를 탐구하는 것은 철학탐구의 좋은 방법이다. 또한 역설을 해소하고 수수께끼를 푸는 과정에서 이성의 힘이 성장한다. 나는 잠자는 미녀 문제의 답이 1/3이라고 오랫동안 강하게 믿고 있다가 이제 갑자기 1/2이라고 생각하게 되었는데 이것은 나에게 몹시도 특이한 경험이다.[27]

1/3주의 해석에서는, 미녀가 깨어났을 때 지금 받고 있는 질문이 첫째 질문일 확률은, N이 무한대로 커질 경우, 0으로 낮아진다. 이것은 N이 무한대가 될 때 $Pr^{o}(A)$가 0으로 낮아질 것을 요구한다. 미녀가 자신이 오직 하나의 물음만 묻게 되는 세계에 들어왔을 리가 없다고 확신하게 되는 것은 나에게 몹시 불합리해 보였다. 이것은 내가 1/3주의를 버리는 계기가 되었다. 나의 계산들이 틀리지 않았다면 다음과 같은 결론을 내릴 수 있다. 첫째, 미녀 자신이 지금 받고 있는 물음이 첫째 물음일 확률은 항상 1/2보다 크다. 둘째, 지금 물음이 첫째 물음이라는 정보는 이미 앞면이 나왔음을 뒷받침하는 단서로 쓰일 수 있다. 셋째, 미녀의 깨어남은 자신이 어떤 세계에 들어와 있는지에 대한 믿음 강도를 변경시키지 않는다. 넷째, 미녀가 처음 깨어나기 이전에 동전을 던지느냐, 아니면 첫째 물음을 마친 뒤에 동전을 던지느냐 하는 것은 확률 계산에서 차이를 낳는다.

---

27) 나의 갑작스런 이 개종이 김한승의 관점주의를 강화해주는지도 모르겠다. 이 점을 보다 깊이 성찰한 후 그의 관점주의를 다룰 기회를 얻었으면 한다.

참고 문헌

김명석 (2009), “우리는 미녀에게 무엇을 물었을까?”, 논리학회 2009년 겨울 학술대회 발표문.

김한승 (2009), “비개념적 내용으로서의 지표적 내용: ‘잠자는 미녀’ 문제에 대한 관점주의적 대답”, 『철학적 분석』 20호, pp. 119-140.

김한승 (2011), “확률에 대한 관점주의”, 『논리연구』 14집 2호, pp. 59-84.

송하석 (2011), “잠자는 미녀의 문제, 그의 대답은”, 『논리연구』 14집 1호, pp. 1-22.

Elga, A. (2000), “Self-locating Belief and the Sleeping Beauty Problem”, *Analysis* **60**, pp. 143-147.

Lewis, D. (2001), “Sleeping Beauty: Reply to Elga”, *Analysis* **61**, pp. 171-176.

Kim, H. (2010), “Sleeping Beauty’s Reflection: In and Out”, 『논리연구』 13집 1호, pp. 21-52.

Kim, N. (2009), “Sleeping Beauty and Shifted Jeffrey Conditionalization”, *Synthese* **168** (2), pp. 295-312.

대안대학원 생각실험실 연구원
Thinking Lab., Seoul
E-mail: myeongseok@gmail.com

When Sleeping Beauty Awaked: An Argument for 1/2

Myeoseok Kim

Some Korean Philosophers has manifested their opinions on Sleeping Beauty problem. For example, Hasuk Song and Namjoong Kim stands for an thirder, while Hanseung Kim for a perspectivistic compatibilist. In order to fill a vacant position, I shall make an argument for halfers in this paper. My presumption is that the probability the question now given to sleeping beauty is the first question among several is greater than thirder's calculated value. Futhermore, I argue that the probability the coin landed heads on condition that the question now given to sleeping beauty is the first question is greater than 1/2.

Key Words: Sleeping Beauty problem, Probability, Elga, Lewis, Hasuk Song, Halfer